Corrigé TD8 + explications certificats

Certificats

Commençons par quelques éclaircissements au sujet de ce qu'est un certificat dans le cadre de la résolution d'un problème de décision. Oubliez pour le moment la définition de problème NP, l'objectif est d'abord de vous faire comprendre ce qu'est un certificat de manière générique. Pour ce faire, je vous propose une analogie.

Imaginez avoir devant vous un tas infini de cadenas. Certains peuvent s'ouvrir à l'aide d'un code à trois chiffres, d'autres ne s'ouvrent pour aucun code. A priori, si je vous donne un cadenas et que je vous demande s'il s'ouvre, c'est long de répondre à la question car il vous faudra potentiellement tester toutes les combinaisons à 3 chiffres pour le savoir. En revanche, si je vous donne un cadenas qui s'ouvre accompagné du code correspondant, il vous sera très facile de vérifier que le cadenas s'ouvre : il suffira de tester le code sur le cadenas.

Dans cette analogie, l'ensemble des cadenas représente l'ensemble des instances d'un problème de décision, les cadenas qui s'ouvrent ses instances positives, les cadenas qui ne s'ouvrent pas ses instances négatives et l'ensemble des codes à trois chiffres est l'ensemble des certificats pour le problème. Plus formellement, si x est une instance positive d'un problème, un certificat pour x consiste en de l'information supplémentaire en plus de l'entrée x permettant de vérifier rapidement que x est bien une instance positive.

Certificats dans le cadre de problèmes NP

Dans le cadre d'un problème NP-complet, on exige que, pour chaque instance positive x de taille n du problème :

1. Il existe un certificat associé à x.
2. La taille de ce certificat est polynomiale en n.
3. Vérifier que ce certificat en est bien un se fait en temps polynomial en n.

Autrement dit, tout ce qu'on fait c'est garantir l'existence d'un certificat pour chaque instance positive et imposer des contraintes dessus. Par exemple, pour le problème SAT, pour toute formule φ satisfiable, il existe une valuation v qui la satisfait ; la taille de v est linéaire en la taille de φ et vérifier que v satisfait effectivement φ se fait en temps linéaire en la taille de φ. Cette valuation v est donc un certificat associé à φ satisfaisant les contraintes 2 et 3 et comme il en existe au moins une pour chacune des formules satisfiables, SAT est NP.

La définition d'un problème NP donnée en cours est strictement équivalente à cette description.

Corrigé exercice 5

Pour mettre en pratique les explications ci-dessus, je propose une correction des questions 5 et 6 avec deux formalismes : celui qui colle à la définition donnée en cours et celui présenté ci-dessus (qui est celui attendu dans une copie).

5) Formulation cours : On pose $C = \{A' \mid A' \subset A\}$, f la fonction identité et B le problème de décision suivant :

$$\begin{cases} \textbf{Entrée} : \text{Un graphe } G = (S, A) \text{ et } A' \subset A. \\ \textbf{Question} : A' \text{ est-il un couplage parfait dans } G \ ? \end{cases}$$

Le problème B est polynomial car pour vérifier que A' est un couplage parfait dans G, il suffit par exemple d'appliquer l'algorithme suivant à G, A' : initialiser un tableau T de $|S|$ cases contenant des zéros (ce tableau est destiné à contenir en case i le nombre d'arêtes de A' qui touchent le sommet i) puis pour chaque arête $(u, v) \in A'$, ajouter un à $T[u]$ et $T[v]$. On renvoie "oui" si T contient des uns dans toutes ses cases et "non" sinon. Cet algorithme s'exécute en temps $O(|S| + |A'|) = O(|S| + |A|) = O(|(G, A')|)$.

On a bien que, pour toute instance $G = (S, A)$ de COUPLAGE PARFAIT, G admet un couplage parfait (c'est-à-dire, G est une instance positive pour COUPLAGE PARFAIT) si et seulement si il existe un élément $A' \in C$ tel que $|A'| \leq |A| \leq |G| = f(|G|)$ et tel que A' est un couplage parfait dans G (c'est-à-dire, (G, A') est une instance positive pour le problème B).

Formulation plus légère : Pour toute instance positive $G = (S, A)$ de COUPLAGE PARFAIT, il existe $A' \subset A$ dont la taille est polynomiale en $|G|$ qui est un couplage parfait de G et vérifier que A' est effectivement un couplage parfait de G se fait en temps polynomial en $|G|$ via l'algorithme décrit ci-dessus.

6) Formulation cours : On pose $C = \{X \mid X \subset S\}$, f la fonction identité et et B le problème de décision suivant :

$$\begin{cases} \textbf{Entrée} : \text{Un graphe } G = (S, A), \text{ un entier } k \text{ et } X \subset S. \\ \textbf{Question} : \text{Y'a-t-il au moins } k \text{ arêtes de } G \text{ entre } X \text{ et } S \setminus X \ ? \end{cases}$$

Le problème B est dans P car voici un algorithme polynomial en $|(G,k,X)|$ renvoyant "oui" si (G,k,X) est une instance positive de B et "non" sinon : pour tout $u,v \in X \times S \setminus X$, on regarde si $(u,v) \in A$ et si oui, on incrémente un compteur. Si ce compteur est supérieur à k à la fin, on renvoie "oui", sinon on renvoie "non". Si on compte extrêmement rigoureusement le coût de cet algorithme :

- Il y a au plus $|S|^2$ couples $(u,v) \in X \times S \setminus X$ (et en fait moins) et pour chacun, vérifier si $(u,v) \in A$ se fait soit en temps constant si G est représenté par matrice d'adjacence soit en temps $O(|S|)$ si il est représenté par listes d'adjacence : cette étape coûte au pire un $O(|S|^3)$.
- Il faut ensuite comparer un nombre inférieur à $|A|$ avec k ce qui se fait en le nombre de bits nécessaire à écrire chacun de ces deux entiers donc en $O(\max(\log|A|, \log k)) \leq O(|A| + \log k)$.

Comme $|S|$, $|A|$ et $\log k$ sont inférieures à la taille de $|(G,k,X)|$, on obtient bien un algorithme polynomial en cette taille. *Remarque : On ne vous demandera a priori jamais d'être aussi précis que ça.*

On a bien que, pour toute instance $(G=(S,A),k)$ de COUPE, (G,k) est une instance positive de COUPE si et seulement si il existe $X \in C$ tel que $|X| \leq |S| \leq f(|G,k|)$ et tel que (G,k,X) est une instance positive du problème B.

Formulation plus légère : Pour toute instance positive $G=(S,A),k$ de COUPE, il existe $X \subset S$ — dont la taille est polynomiale en la taille de cette instance — tel qu'il y a au moins k arêtes entre X et $S \setminus X$ dans G. Vérifier qu'un tel X montre que (G,k) est une instance positive de COUPE peut se faire en temps polynomial en la taille de (G,k) en parcourant tous les couples (u,v) de $X \times S \setminus X$ et en vérifiant que pour au moins k d'entre eux, $(u,v) \in A$.

Corrigé exercice 6

On note CH = CYCLE HAMILTONIEN et PVC = VOYAGEUR DE COMMERCE.

1. Vérifier qu'une suite C de moins de $|S|$ sommets est un cycle hamiltonien dans le graphe $G=(S,A)$ peut se faire en temps polynomial en $|G|$ en vérifiant qu'il n'y a pas de doublons dans C, que $|C|=|S|$ et qu'il y a bien une arête de G entre deux sommets consécutifs de C (y compris entre le dernier et le premier). Donc CH $\in$ NP.

   Montrons maintenant que CHO $\leq$ CH. Soit $G=(S,A)$ un graphe orienté. On construit le graphe $G'=(S',A')$ comme suit (a priori, cette construction vous serait donnée dans un devoir car elle n'est pas facile à inventer) :

   - $S' = \bigcup_{s \in S} \{s_1, s_2, s_3\}$: autrement dit, on détriple chaque sommet de G.
   - $A' = \bigcup_{s \in S} \{(s_1,s_2),(s_2,s_3)\} \cup \{(s_3,t_1) \,|\, (s,t) \in A\}$, ces arêtes étant non orientées.

   Le graphe G' est non orienté donc est une instance pour CH. Il est possible de construire G' en temps polynomial en $|G|$ car il y a $3|S| = O(|G|)$ sommets et $|A|+2|S| = O(|G|)$ arêtes dans G'. Enfin, G est une instance positive de CHO si et seulement si G' est une instance positive de CH. En effet :

   ($\Rightarrow$) Si G admet un cycle hamiltonien $s^{(1)}, s^{(2)}, ..., s^{(n)}$, alors $s_1^{(1)}, s_2^{(1)}, s_3^{(1)}, s_1^{(2)}, s_2^{(2)}, s_3^{(2)}, ..., s_3^{(n)}$ est un cycle hamiltonien dans G' (en fait, chaque sommet se transforme en trois sommets numérotés par $1,2,3$ et il suffit de parcourir ces 3 sommets dans le sens $1 \to 2 \to 3$).

   ($\Leftarrow$) Réciproquement, si G' admet un cycle hamiltonien C alors comme C n'est pas orienté, on peut supposer que ce cycle parcourt les sommets s_1, s_2, s_3 dans le sens $1 \to 2 \to 3$ pour tout sommet $s \in S$. Ceci donne naissance à un cycle dans G par "fusion" pour tout $s \in S$ des trois sommets (nécessairement consécutifs dans C) s_i en le sommet s.

   On vient donc d'établir que CHO se réduit polynomialement en CH et comme CHO est NP-difficile, CH aussi. Ce problème est NP-difficile et NP donc est NP-complet.

2. La preuve du caractère NP de PVC se fait de manière similaire à celle de CH.

   Soit $G=(S,A)$ une instance de CH. Alors le graphe $G'=(S,A)$ pondéré par la fonction p qui donne comme poids 1 à toutes les arêtes est un graphe qu'on peut construire polynomialement en $|G|$ donc $(G',|S|)$ est une instance de PVC constructible en temps polynomial en $|G|$.

   Or, si G admet un cycle hamiltonien, alors ce cycle est un cycle hamiltonien dans G' de poids égal à $|S|$ et réciproquement, si G' admet un cycle hamiltonien de poids supérieur ou égal à $|S|$ alors ce cycle est aussi un cycle hamiltonien dans G. Donc G est une instance positive pour CH si et seulement si G' est une instance positive de PVC.

   Ceci montre que CH $\leq$ PVC et comme CH est NP-difficile d'après la question 1, PVC aussi ce qui était le point qui manquait pour montrer sa NP-complétude.